

**PRINTING SYSTEM WITH EMBEDDED AUDIO/VIDEO CONTENT  
RECOGNITION AND PROCESSING**

Inventors:  
Jonathan J. Hull  
Berna Erol  
Jamey Graham  
Peter E. Hart  
Dar-Shyang Lee  
Kurt Piersol  
Daniel G. Van Olst

**CROSS REFERENCE TO RELATED APPLICATIONS**

**[0001]** This application claims the benefit of the following provisional patent applications, each of which is incorporated by reference in its entirety: U.S. Provisional Application No. 60/506,303, filed September 25, 2003; and U.S. Provisional Application No. 60/506,302, filed September 25, 2003.

**[0002]** This application is a continuation-in-part of the following co-pending patent applications (hereinafter, "the Video Paper patent applications"), each of which is each incorporated by reference in its entirety: U.S. Application No. 10/001,895, filed November 19, 2001; U.S. Application No. 10/001,849, filed November 19, 2001; U.S. Application No. 10/001,893, filed November 19, 2001; U.S. Application No. 10/001,894, filed November 19, 2001; U.S. Application No. 10/001,891, filed November 19, 2001; U.S. Application No. 10/175,540, filed June 18, 2002; and U.S. Application No. 10/645,821, filed August 20, 2003.

**[0003]** This application is also related to the following co-pending patent applications, each of which is incorporated by reference: U.S. patent application

entitled, "Printer Having Embedded Functionality for Printing Time-Based Media," to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08340; U.S. patent application entitled, "Networked Printing System Having Embedded Functionality for Printing Time-Based Media," to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08341; U.S. patent application entitled, "Stand Alone Multimedia Printer Capable of Sharing Media Processing Tasks," to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08342; U.S. patent application entitled, "Printer With Hardware and Software Interfaces for Peripheral Devices," to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08383; U.S. patent application entitled, "Networked Printer With Hardware and Software Interfaces for Peripheral Devices," to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08384; U.S. patent application entitled, "Stand Alone Printer With Hardware / Software Interfaces for Sharing Multimedia Processing," to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08385; U.S. patent application entitled, "Printer User Interface," to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08455; U.S. patent application entitled, "User Interface for Networked Printer," to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08456; and U.S. patent application entitled, "Stand Alone Multimedia Printer With User Interface for Allocating Processing," to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08457.

## **BACKGROUND**

### **Field of the Invention**

**[0004]** This invention relates generally to media content recognition and processing, and in particular to printing systems having embedded logic for audio and/or video content recognition and processing that can generate a printed representation for the audio and/or video content.

### **Background of the Invention**

**[0005]** A conventional printer can receive documents or other data in a number of formats and then prints the contents of those documents or data in accordance with the proper format. But while conventional printers can print documents in a wide variety of formats, these printers are fundamentally limited in their ability to reproduce different kinds of media. For example, it is standard technology for a printer to produce images of static text, pictures, or a combination of the two. But because these printers print onto paper or another similar fixed medium, they cannot record the nuances of time-based media very well.

**[0006]** Accordingly, existing printers are not designed to generate multimedia documents, and there is no effective method for generating an easily readable representation of media content in any kind of printed format. Several different techniques and tools are available for accessing and navigating multimedia information (e.g., existing media renderers, such as Windows Media Player); however, none of these provide the user with the option of creating a multimedia document that the user can easily review and through which a user can gain access to media content.

**[0007]** There are many recognition and processing software applications that can be applied to audio or video content, for example, face recognition, scene detection, voice recognition, etc. But the limitations of existing printing systems described above reduce the utility of these applications. Without a paper-based or other printed representation of the processed media, the utility of these applications remains in the electronic domain. This is because the current state of the art requires a user to install and maintain these applications on a computer, which can only display the results electronically. Moreover, these applications often require significant resources of the computer, such as memory and processor speed, thus inhibiting their widespread use.

**[0008]** What is needed therefore is a printing system that is equipped to print time-based media without the limitations of conventional printers. It is further desirable that such a printer be able to perform at least some of the necessary processing itself rather than require an attached computer or other device to perform all of the processing.

## **SUMMARY OF THE INVENTION**

**[0009]** To overcome at least some of the limitations of existing printing systems, a printing system in accordance with an embodiment of the invention includes embedded hardware and/or software modules for performing audio and/or video content recognition and processing. In addition, the printing system can generate a paper-based or other printed representation based on the results of the content recognition and processing performed on the audio and/or video content. In this way, a user can obtain a useful printed result of time-based media content based on any number of different processing needs. Moreover, packaging these capabilities on the printer relieves the resource burden on another device, such as an attached computer or a source device.

**[0010]** In one embodiment, a printer receives time-based media data that includes audio and/or video data. Using embedded software and/or hardware modules, the printer segments the data according to a content recognition and processing algorithm. The results of this algorithm may include one or more of: data for producing a printed representation of the media data, meta data corresponding to the segmentation of the media data, and an electronic representation of the media data. The printer then produces a printed output based on the segmentation of the media data, the printed output including for example samples of the media content where the content was segmented as well as information related to those samples. Using the printed representation of the time-based media, the user can quickly view and access the media at desired places therein. The printer may also generate an electronic version of the media data, which may be identical to the received data or modified.

**[0011]** The printer's embedded content recognition and processing functionality can perform a variety of functions depending on the desired application for the printer.

Without intending to limit the types of processing functions, in some embodiments the printer includes embedded modules for providing at least a portion of the processing for one or more of the following functionalities: video event detection, video foreground/background segmentation, face detection, face image matching, face recognition, face cataloging, video text localization, video optical character recognition, language translation, frame classification, clip classification, image stitching, audio reformatting, speech recognition, audio event detection, audio waveform matching, audio-caption alignment, caption alignment, and any combination thereof.

**[0012]** In one embodiment, the meta data produced from the media data are embedded within the printed representation, such as in a bar code next to a sample. In another embodiment, the printer generates an electronic version of the media data that includes the meta data, which contain the segmentation information.

**[0013]** In another embodiment, a system for printing time-based media data includes a media renderer for viewing a selected media item, where the media renderer includes a print option. When a user selects the print function for a viewed media item, a printer driver sends the media item to a printer. The printer then segments the media item according to a content recognition algorithm and produces a printed output based on the segmented media item. The printed output includes a plurality of samples of the media item and information related to the samples. In this way, a media renderer can be equipped with a print function. In one embodiment, a plug-in module for a standard media renderer (e.g., Windows Media Player and Real Media Player) provides the print

function, thus providing a print functionality for existing widely used renderers that currently do not have that capability. Once the print function is selected, the user can interact with the content recognition modules on the printer to create a printed representation of the media that represents the recognition routines that were applied to the selected media.

**[0014]** In addition to relieving external devices of the computation load required by various content recognition and processing algorithms, embedding these functionalities in the printer may allow for multiplatform functionality. Embedding functionalities within a printer also lead to greater compatibility among various systems, and it allows content recognition and processing in a printer that acts as a walk-up device in which no attached computer or other computing system is required.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

**[0015]** FIG. 1 is a diagram of an overview of an embodiment of a printing system with embedded audio and/or video content recognition and processing capabilities, in accordance with an embodiment of the invention.

**[0016]** FIG. 2 shows the process flow of the operation of a printing system with embedded audio and/or video content recognition and processing capabilities, in accordance with an embodiment of the invention.

**[0017]** FIG. 3 shows an example of meta data that may be created for a segmentation, in accordance with an embodiment of the invention.

**[0018]** FIG. 4 is a diagram of one embodiment of a printing system coupled to an attached computing device equipped with a media rendering application, in accordance with an embodiment of the invention.

**[0019]** FIG. 5 is a flow diagram of the operation of the printing system and media renderer application, in accordance with an embodiment of the invention.

**[0020]** FIG. 6 is a screen shot of one embodiment of the media renderer application.

**[0021]** FIG. 7 is a screen shot of one embodiment of a print dialog window.

**[0022]** FIG. 8 is a screen shot of one embodiment of a print dialog window showing a preview of the printed output.



## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0023] Various embodiments of a printing system include embedded functionality for performing content recognition algorithms on received media content. In this way, the printing systems can perform content-based functionalities on time-based media and then print the results of these operations in a useful and intelligent format. Depending on the desired application, the printing system may perform any among a number of content recognition and processing algorithms on the received media content. Moreover, the printing system may include any number of devices for receiving the media, printing the printed output, and producing the electronic output. Therefore, a number of embodiments of the printing system are described herein to show how such a system can be configured in a virtually limitless number of combinations to solve or address a great number of needs that exist.

### System Overview

[0024] FIG. 1 illustrates a multifunction printer 100 having embedded functionality in accordance with an embodiment of the invention. As FIG. 1 illustrates, the printer 100 can receive media data, which may include audio data, video data, or a combination thereof. The printer 100 includes embedded functional modules 105 for performing content recognition and processing and user interaction. The printer 100 may also use the modules 105 to create a printed output 110 and associated media data 120. The printer 100 can also communicate with a user through an external device, such as a computer system or other electronic device that can communicate commands and data with the printer 100. This interactive communication allows a user to interact with the embedded functionality within the printer 100, for example to provide commands to

control how the printer 100 performs the content recognition and processing on the received media data.

[0025] Depending on the desired application, the functional modules 105 may perform any number of content recognition and processing algorithms, including video event detection, video foreground/background segmentation, face detection, face image matching, face recognition, face cataloging, video text localization, video optical character recognition, language translation, frame classification, clip classification, image stitching, audio reformatting, speech recognition, audio event detection, audio waveform matching, audio-caption alignment, caption alignment, and any combination thereof.

[0026] In one embodiment, the printer 100 is a multifunction printer as described in co-pending U.S. patent application entitled, "Printer Having Embedded Functionality for Printing Time-Based Media," to Hart et al., filed March 29, 2004, Attorney Docket No. 20412-08340, which application is incorporated by reference in its entirety; a networked multifunction printer as described in co-pending U.S. patent application entitled, "Networked Printing System Having Embedded Functionality for Printing Time-Based Media," to Hart et al., filed March 29, 2004, Attorney Docket No. 20412-08341, which application is incorporated by reference in its entirety; or a stand-alone multifunction printing system as described in co-pending U.S. patent application entitled, "Stand Alone Multimedia Printer Capable of Sharing Media Processing Tasks," to Hart et al., filed March 29, 2004, Attorney Docket No. 20412-08342, which application is incorporated by reference in its entirety.

[0027] The printer 100 may receive the audio and/or video data from any of a number of sources, including a computer directly, a computer system via a network, a portable device with media storage (e.g., a video camera), a media broadcast to an embedded media receiver, or any of a number of different sources. Depending on the source, the printer 100 includes appropriate hardware and software interfaces for communicating therewith, such as the embodiments described in co-pending U.S. patent application entitled, "Printer With Hardware and Software Interfaces for Peripheral Devices," to Hart et al., filed March 29, 2004, Attorney Docket No. 20412-08383; co-pending U.S. patent application entitled, "Networked Printer With Hardware and Software Interfaces for Peripheral Devices," to Hart et al., filed March 29, 2004, Attorney Docket No. 20412-08384; and co-pending U.S. patent application entitled, "Stand Alone Printer With Hardware / Software Interfaces for Sharing Multimedia Processing," to Hart et al., filed March 29, 2004, Attorney Docket No. 20412-08385; all of which are incorporated by reference in their entirety.

[0028] Moreover, the interactive communication can be provided by a user interface in the form of a display system, software for communicating with an attached display, or any number of embodiments as described in co-pending U.S. patent application entitled, "Printer User Interface," to Hart et al., filed March 29, 2004, Attorney Docket No. 20412-08455; co-pending U.S. patent application entitled, "User Interface for Networked Printer," to Hart et al., filed March 29, 2004, Attorney Docket No. 20412-08456; and co-pending U.S. patent application entitled, "Stand Alone Multimedia Printer With User Interface for Allocating Processing," to Hart et al., filed March 29,

2004, Attorney Docket No. 20412-08457; all of which are incorporated by reference in their entirety.

[0029] FIG. 2 illustrates a typical operational flow of one embodiment of the printer 100. To obtain a printed output of media data based on a content recognition and processing algorithm, a user transfers 205 the media data to the printer 100. As described above, the transferring 205 can be performed through any of a number of interfaces. The printer 100 then performs 210 a desired content recognition algorithm on the transferred media data. Which algorithm is performed, as well as parameters for the algorithm, may be selected by the user during interactive communication with the printer 100, as described in greater detail below. The result of the algorithm is to segment the time-based media data, which represents audio and/or video content. Segmenting the media data may result in a number of samples of the media data at various time locations in the data. Alternatively, segmenting the media data may result in a set of ranges, a set of start frames or times, a set of end frames or times, or any other set of data that results in a division or organization of the time-based media content.

[0030] Having segmented the media data, the printer 100 generates 215 meta data that describes the segmentation. In this way, the meta data can be associated with the segmented media data to indicate the location of particular samples of the media data within that content. The meta data may further include information about the segments or samples that define the segmentation. For example, the printer may employ a content recognition algorithm, such as a facial recognition algorithm, on a particular frame of data associated with a segment. The meta data could then include the result of the content recognition algorithm, such as the identity of the person recognized by the

algorithm. The meta data may further include information for associating the segments or samples of the media data with their occurrence in the media data, using for example time stamps.

**[0031]** The printer then produces 220 a printed output 110 of the media data based on the results of the content recognition algorithm. The printed output 110 may include a representation of a sample from the media data as well as information obtained using the content recognition algorithm, which may describe the sample or associated segment. For example, the printed output 110 may include a number of entries, each of which contains an image of a face from a video data input, a name of a person associated with that face using a facial recognition algorithm, and other data such as a time stamp for when the face appeared in the video. The printer 100 may also encode information on the printer output 110, for example on a bar code, which includes information or indicia relating to the segment. In one embodiment, the printed output 110 is video paper, as described in the Video Paper patent applications, referenced above.

**[0032]** In one embodiment, the printer 100 also produces 225 an electronic representation of the media data 120, which representation may be identical to the received data, a reformatted version of the received data, or a modified version of the received data. Rather than being included on the printed output 110, the meta data that were generated 215 may be encoded entirely or in part within the electronic representation of the media data 120. In another embodiment, the media data are available by other means, so the printer 105 need not generate an additional media data output 120.

[0033] In this way, the printer 100 can be used to print time-based media data to create a useful and intelligent representation of a time-based even on a two-dimensional output. The content recognition algorithms can be selected to segment the time-based media data and to retrieve information from the data. The resulting segmentation and retrieved information are represented in a useful way on the printed output 110. By linking the printed output 110 with the media data output 120, for example using the meta data, the information presented in the printed output 110 can be easily associated with its actual occurrence. FIG. 3 illustrates an example of meta data used to link a segmented video file with its electronic representation. This enables a number of different embodiments for intelligent printers, described in greater detail below.

#### Printing Media from a Media Renderer

[0034] In one application, the printer 100 is used to create a printed representation of media data that is viewed on a computer. FIG. 4 shows an example environment in which a printer 100 is connected to a computer system 340 for printing media data. As shown, the printer 100 includes an interface 310 for communicating with the computer 340, an output system for producing printed and/or electronic representations of the media, and embedded functional modules 105. In the example shown, the printer 100 includes a user interaction module 320, a content recognition module 330, and a printed output generation module 335. The user interaction module 320 comprises hardware and/or software that allows the printer 100 to communicate control signals and media data with the computer 340 through the interface 310. The user interaction module 320 also allows the user to interact with the printer through the attached computer 340. The content recognition module 330 comprises hardware and/or software for performing the

content recognition and processing of the media data, including the segmentation of the media data and the creation of the meta data. The printed output generation module 335 comprises hardware and/or software for generating the desired printed output using the output system 315 based on the results of the content recognition and segmentation.

The output system 315 includes hardware for writing media data, such as a DVD writer, a secure digital writer, a network interface card, or any other suitable device for providing the media data output.

**[0035]** As described, the content recognition module 330 may perform any number of algorithms on the media data, including video event detection, video foreground/background segmentation, face detection, face image matching, face recognition, face cataloging, video text localization, video optical character recognition, language translation, frame classification, clip classification, image stitching, audio reformatting, speech recognition, audio event detection, audio waveform matching, audio-caption alignment, and caption alignment.

**[0036]** In this example, the computer 340 includes media data 350 storage for example on a storage device within or in communication with the computer 340. Installed on the computer 340 is a printer driver 345, which allows the computer 340 to communicate with the printer 100, including sending media data to be printed and print commands in a predefined printer language. In this embodiment, the computer 340 also includes a media rendering application 355, such as Windows Media Player or Real Media Player. Using the media rendering application 355, a user can play back media data on the computer, such as viewing video files and listening to audio files. The media rendering application 355 further includes a "print" function, which a user can select to

initiate printing a currently viewed or opened media item, such as a video clip. Upon invocation of the print function, the printer driver 345 transfers the media data to the printer 100, instructs the printer 100 to apply one or more content recognition algorithms, and provides any appropriate parameters for those algorithms.

[0037] FIG. 5 illustrates one process for printing a media item using a media rendering application 355, such as that shown in the screen shot of FIG. 6. To initiate the printing process, a user presses 505 a print button or otherwise invokes the print function in the media rendering application 355. In response, the application 355 displays 510 a print dialog, such as that shown in FIG. 7. The print dialog allows the user to select 515 parameters for generation of the printed output. Although FIG. 7 is merely an example for a particular application, typical parameters that can be entered in the print dialog include a printer destination, document formatting options (including the types of information to be displayed on the printed output), parameters that affect the content recognition algorithm, and a destination for an electronic version of the media data.

[0038] Once the user selects 515 the desired parameters and approves 520 them by selecting an update or a print function, the printer driver 345 sends 525 the parameters to the printer 100. The update function is for directing the printer 100 to perform the desired processing and return a preview of the output to the printer dialog, while the print function is for directing the printer 100 to perform the desired processing and actually produce a corresponding output. If 530 the media data are not already transferred to the printer 100, the driver 345 sends 535 the media data to the printer 100 as well. With the media data and the parameters for transforming the media data known,



the printer 100 can determine an appropriate printed output for the media data. If 540 this processing has not been completed, the printer performs 545 the requested function.

[0039] If 550 the user had selected the update function, the printer 100 returns 560 the processed data to the media rendering application 355 via the printer driver 345, and the media rendering application updates a preview of the output, as shown in FIG. 8.

Otherwise, if 550 the user had selected the print function, the printer 100 generates 555 the requested printed output and electronic output, if any, according to the parameters and data from the print dialog. This step may include assignment of an identifier to the media data to link the printed and media data outputs. In this way, the user has printed a media item view from a media player, where the printed output represents the media data based on a selected content recognition functionality.

[0040] In the past, media renderers did not have a print function because, without the printer 100 described herein, there was no way for a user to generate a meaningful printout for an arbitrary video or audio file. With the printer 100 having embedded functional modules 105, as described herein, techniques for transforming media into two-dimensional representations inside the printer are provided. It thus becomes useful for a media renderer to have a print function, similar to a word processor or any other application that opens documents.

[0041] In one embodiment, the print functionality is provided by a plug-in module 360, which allows a standard media renderer to take advantage of the printing capabilities of the multifunction printer 100. For example, a print option can be added to the Windows Media Player (WMP) version 9 using the plug-in feature provided by Microsoft. The plug-in feature allows developers to create an application that

supplements the WMP in some way. For instance, someone might write code that displays a graphic equalizer inside the WMP to show the frequency distribution for a particular audio track or the audio from a video file. Microsoft provides an explanation of what a plug-in is and how to build a plug-in at: "Building Windows Media Player and Windows Media Encoder Plug-ins" by David Wrede, dated November 2002, which can be accessed through the Microsoft developer website at *msdn.microsoft.com*. As explained, several types of plug-in can be created, such as display, settings, metadata, window, and background. Using one of the user interface (UI) plug-in styles, a button or panel can be added to the WMP screen. If a button were added, for example, depending on the type of plug-in chosen, the button would be located in a specific area in the WMP's display window. The plug-in module 360 could thus be bundled and registered as a dynamically linked library (DLL), and the computer code for performing the desired action could be included in the DLL or invoked by the DLL when the button is pressed. In another embodiment, a print option is added to the File menu of WMP, using the "hooking" technique described in the Wrede article. Although this technique may be slightly less elegant than a plug-in, it would put a print option where it normally appears in most other document rendering applications.

[0042] FIG. 6 shows an example of a standard media player having a print function enabled by a plug-in module 360. To create an application that generates a printed representation from a given video file, the plug-in module 360 (e.g., in the form of a DLL) of the *metadata* type is created and a print button is placed on the right side of the video pane of the WMP application. Also installed on the computer 340 is a Video Paper application module (not shown), as described in the Video Paper patent

applications, referenced above. The plug-in module 360 is programmed to invoke the Java application and pass it the necessary parameters that specify the layout of the video paper document. Typically, these include the number of key frames per page, their size, placement, and if any text is present (perhaps in a closed caption), how it should be formatted (e.g., point size, line length, justification, etc.). For example, if the video file currently loaded in the media rendering application 355 were from the media database 350 and it was in the MuVIE format (i.e., it includes metadata such as xml files and keyframes), then the invocation would pass an ID for the video file (e.g., 00043) plus a path name (d:/media) to the Video Paper application module along with default parameters. The Video Paper application module would create the video paper, for example in PDF form, and the video paper would be output 315 in electronic 120 or paper 110 form.

**[0043]** As explained, the print driver 345 allows for interactive communication between a user operating the computer 340 and the printer 100. Printer drivers are normally not designed to facilitate interactive information gathering. Because the print job can be redirected to another printer, or because printing protocols do not typically allow such interactive sessions, operating systems generally discourage interaction with the user by a print driver. Once initial printer settings are captured, further interactions are generally not allowed. One way to add this ability to a print driver is to embed metadata into the print stream itself. However, it is possible that the printer could need to ask the user for more information, in response to computations made from the data supplied by the user. In addition, the printer might itself delegate some tasks to other application servers, which might in turn need more information from the user. So-called

“Web services” or “grid computing” systems are examples of the sort of application server that the printer might trigger.

**[0044]** In order to allow this interaction, without modifying printer driver architecture of the underlying operating system, an extra mechanism called a UI Listener is constructed. The UI Listener, a program that listens to a network socket, accepts requests for information, interacts with a user to obtain such data, and then sends the data back to the requester. Such a program might have a fixed set of possible interactions or accept a flexible command syntax that would allow the requester to display many different requests. An example of such a command syntax is a standard web browser’s ability to display HTML forms. These forms are generated by a remote server and displayed by the browser, which then returns results to the server. A UI listener is different from a browser, though, in that a user does not generate the initial request to see a form. Instead, the remote machine generates this request. The UI listener is a server, not a client.

**[0045]** Because network transactions of this type are prone to many complex error conditions, a system of timeouts are used to assure robust operation. Normally, each message sent across a network either expects a reply or is a one-way message. Messages which expect replies generally have a timeout, a limited period of time during which it is acceptable for the reply to arrive. In this invention, embedded metadata would include metadata about a UI listener that will accept requests for further information. Such metadata consists of at least a network address, port number, and a timeout period. It might also include authentication information, designed to prevent malicious attempts to elicit information from the user. Since the user cannot tell whether the request is coming

from a printer, a delegated server, or a malicious agent, prudence suggests strong authentication by the UI listener. If the printer or a delegated application server wishes more information, it can use the above noted information to request that the UI listener ask a user for the needed information.

#### Additional Applications for a Printer With Embedded Content Recognition Functionality

[0046] In addition to the embodiments described, the multifunction printer 100 can be applied in many other configurations to achieve a variety of results. To illustrate the wide variety of uses for a printer having embedded content recognition functionality, a number of additional embodiments and applications for the printer are described. These embodiments are described to show the broad applicability for such a printer and are therefore not meant to limit the possible applications or uses for the printer.

##### Printer with embedded video event detection

[0047] When the user prints a video, a set of events (e.g., camera motion) are detected and used to generate a Video Paper document that provides index points to those events in the video. The document could also provide symbolic labels for each event. For example, “camera swipe, left-to-right, at 00:12:52.”

##### Printer with embedded video foreground/background segmentation

[0048] A printer with a video camera attached includes software for foreground/background segmentation. The printer monitors the appearance of people in the field of view and constructs a video or still-image record of people who walk up to the printer or pass by it. On a personal desktop printer, this system could learn what its owner looks like and store only a limited number of shots of that person (once per day, for example, to show what that person was wearing that day), and store images of the

visitors to the office. Those images could be printed immediately, negating the need for on-printer storage, or they could be queued and formatted for printing later.

Printer with embedded face image detection

[0049] The user prints a JPEG or a video file and face image detection software frames the faces it detects. Software on the client device (the print dialog box) allows the user to print a zoomed-up version of a face image.

Printer with embedded face image matching

[0050] Every still image or video a user prints is subjected to face image extraction and matching against a database resident on the printer that's updated periodically by downloading from a central server. When a match is found, an alert is generated by email, over a speaker attached to the printer, or by refusing to print that document. This technology could be used by a photo lab to scan all the snapshots they print automatically, e.g., to look for terrorists.

Printer with embedded face recognition

[0051] The user prints a video file, and the printer recognizes the face images it contains. A paper printout is provided that shows images of those faces, the symbolic recognition results, and where the face occurred in the video. This will substantially reduce the time needed for someone searching a video file for the instance of a particular individual. With this embodiment, a person can quickly scan a paper document rather than watching a recording.

#### Printer with embedded face extraction, matching, and cataloging

[0052] A user prints a video or still picture file, and the face images it contains are extracted and stored on the printer. Subsequently, the printer monitors other video or still image files and, as they are printed, attempts to match the face images they contain to the database. From the print dialog box, the user can preview the face extraction results and cross index the face images in a given video to the videos that were printed before the present one. Special cover sheets can be provided to show the faces contained in a given video and the results of cross-indexing.

#### Printer with embedded video text localization

[0053] A user prints a video recording and the locations of all the text in the video are determined. This helps segment the video into scenes by changing text layouts. A cover sheet includes at least one frame from each such scene and lets a user browse through the video and see what text was captured. Printed time stamps or bar codes provide a method for randomly accessing the video. An example use would be printing a home video recording that contains somewhere within it a shot showing the storefront of a leather jacket shop in Barcelona. The user's attention would immediately be drawn to the point in the video containing this information, eliminating the need to watch more than an hour of video to find that point. Note that the reliability of video text localization can be much higher than with optical character recognition (OCR).

#### Printer with embedded video OCR

[0054] The user prints a video file and the text it contains is automatically recognized with an OCR algorithm. A paper printout can be generated that contains only the text or key frames selected from the video plus the text. This provides a

conveniently browsable format that lets a user decide whether to watch a video recording.

Printer with embedded video text foreign language translation

[0055] The user prints a video file, which is then scanned with an OCR algorithm. The recognition results are translated into a foreign language and printed on a paper document together with key frames extracted from the video. The user can follow along while the video is playing and consult the paper document whenever necessary.

Printer with embedded video frame classification

[0056] A user prints a video and the printer classifies each frame into a number of known categories, such as "one person," "two people," "car," "cathedral," "tree," etc. These categories are printed next to each frame on a paper representation for the video. They can also be used, under control of a print dialog box, to generate a histogram of categories for the video that can be printed (like a MuVIE channel) on the printout. This lets a user browse the printout and locate, for example, the section of the home video recording that shows the cathedral in Barcelona.

Printer with embedded video clip classification

[0057] A user prints a video and the printer segments it into scenes and classifies each scene into a number of known categories, such as for example a group interview or a field report. The printout shows a representative key frame from each clip as well as the recognition result and a time stamp or bar code that provides a means for randomly accessing the video. In one example, this lets a user easily find the discussion among five news commentators that occurs sporadically on Fox News.



#### Printer with embedded trainable video clip classification

[0058] A user prints a video and the printer segments it into scenes and classifies each one into a number of known categories. The user is presented a dialog box that shows the result of that classification and allows the user to manually classify each clip. The printer's clip classifier is updated with this information. The printout shows a representative key frame from each clip as well as the original recognition result, the manually assigned category, and a time stamp or bar code that provide a means for randomly accessing each clip.

#### Printer with embedded digital image stitching

[0059] The user prints a set of digital images that are intended for stitching. Under control of a print dialog box, these images are laid out horizontally, vertically, and transformed so that the final printed image has minimal distortion.

#### Printer with embedded audio re-formatter

[0060] The printer includes WAV to MP3 conversion hardware (and/or software). The user prints a WAV file, and a Video Paper document is output as well as an alternative version of the audio file (e.g., MP3 format) that can be played on a client device.

#### Printer with embedded speech recognition

[0061] The user prints an audio file, which is passed through a speech recognition program. The recognized text is printed on a paper document. A representation is provided that indexes the words or phrases that were recognized with high confidence.

The print dialog box provides controls for modifying recognition thresholds and layout parameters.

Printer with embedded audio event detection

[0062] The user prints an audio file, and a set of events (e.g., shouting) are detected and used to generate a Video Paper document that provides index points to those events in the audio. The document could also provide symbolic labels for each event, for example, “loud shouting occurred at 00:12:52.”

Printer with embedded audio waveform matching

[0063] A user prints an audio file. The printer uses a music-matching algorithm to find other recordings of the same piece. The user can choose which recording to print with the print dialog box. The result is a video paper printout, including a digital medium. The dialog box is another way to deliver music matching as a network service. If the client computer also has a microphone, the user could whistle the tune to the printer and it could find a professional recording.

Printer with embedded audio foreign language translation

[0064] The user prints an audio or video file. The audio in the file is passed through speech recognition, and the results are automatically translated into another foreign language. A paper document is generated that shows the translated output.

Printer with embedded audio – caption alignment

[0065] The user prints an audio file and a text transcript of the audio file that is not aligned with the audio in the audio file. The printer aligns the two streams and prints a

video paper version of the transcript. A symbolic version of the alignment result is printed on the document or returned digitally.

Printer with embedded video OCR and caption matching

[0066] The user prints a video recording that includes a closed caption. The text in the video is recognized with an OCR algorithm, and the text that occurs in both the video and the closed caption is used as a cue for key frame selection. Key frames nearby those events are printed on a paper document together with a highlighted form of the text that occurred in both channels.

Printer with embedded closed caption extraction and reformatting

[0067] The user prints a video file, and the closed caption is extracted from the file and reformatted on a paper document together with key frames extracted from the video. This lets a user browse the recording and read what was said, thus substantially improving the efficiency of someone who needs to review hours of video.

Printer with embedded TV news segmentation and formatting

[0068] A user prints a TV news program. Because of the specialized format of a typical news program, the printer can apply special video segmentation and person identification routines to the video. The transcript can be formatted more like a newspaper with embedded headlines that make it easy for someone to browse the paper document.

Printer with embedded audio book speech recognition and formatting software

[0069] The user prints an audio book recording. Because the original data file contains a limited number of speakers, the speech recognition software is trained across

the file first. The recognition results can be formatted to appear like a book, taking into account the dialog that that occurs, and printed on an output document. This may be useful for people who have the tape but not the original book.

#### Printer with embedded audio book foreign language translation and formatting

[0070] The specialized audio book recognition system is applied first, as described above, and the results are input to translation software before layout and printing in a specialized format.

#### Route Planning and Mapping

[0071] In a printer with embedded map generation software for routing, the user enters an address on a print dialog box. The printer then generates a map that shows the location of that address.

[0072] In a printer with embedded route planning, the user enters two addresses on a print dialog box. The printer calculates a route between them (e.g., using A\*). A multi-page map format is then generated, improving upon the standard computer-generated map from the Internet.

#### General Comments

[0073] While examples of suitable printing systems are described above, the description of the printer and its document production means is not meant to be limiting. Depending on the intended application, a printer can take many different forms other than the typical office or home-use printer with which most people are familiar. Therefore, it should be understood that the definition of a printer includes any device that is capable of producing an image, words, or any other markings on a surface.

Although printing on paper is discussed above, it should be understood that a printer in accordance with various embodiments of the present invention could produce an image, words, or other markings onto a variety of tangible media, such as transparency sheets for overhead projectors, film, slides, canvass, glass, stickers, or any other medium that accepts such markings.

**[0074]** In addition, the description and use of media and media data are not meant to be limiting, as media include any information, tangible or intangible, used to represent any kind of media or multimedia content, such as all or part of an audio and/or video file, a data stream having media content, or a transmission of media content. Media may include one or a combination of audio (including music, radio broadcasts, recordings, advertisements, etc.), video (including movies, video clips, television broadcasts, advertisements, etc.), software (including video games, multimedia programs, graphics software, etc.), and pictures (including still images in jpeg, gif, tif, jpeg2000, pdf, and other still image formats); however, this listing is not exhaustive. Furthermore, media and media data may further include anything that itself comprises media or media data, in whole or in part, and media data includes data that describes a real-world event.

Media data can be encoded using any encoding technology, such as MPEG in the case of video and MP3 in the case of audio. They may also be encrypted to protect their content using an encryption algorithm, such as DES, triple DES, or any other suitable encryption technique.

**[0075]** Moreover, any of the steps, operations, or processes described herein can be performed or implemented with one or more software modules or hardware modules, alone or in combination with other devices. It should further be understood that portions

of the printer described in terms of hardware elements may be implemented with software, and that software elements may be implemented with hardware, such as hard-coded into a dedicated circuit. In one embodiment, a software module is implemented with a computer program product comprising a computer-readable medium containing computer program code, which can be executed by a computer processor for performing the steps, operations, or processes described herein.

[0076] In alternative embodiments, the printer can use multiple application servers, acting in cooperation. Any of the requests or messages sent or received by the printer can be sent across a network, using local cables such as IEEE1394, Universal Serial Bus, using wireless networks such as IEEE 802.11 or IEEE 802.15 networks, or in any combination of the above.

[0077] The foregoing description of the embodiments of the invention has been presented for the purpose of illustration; it is not intended to be exhaustive or to limit the invention to the precise forms disclosed. Persons skilled in the relevant art can appreciate that many modifications and variations are possible in light of the above teachings. It is therefore intended that the scope of the invention be limited not by this detailed description, but rather by the claims appended hereto.